countries were the following: (1) keep the same date, event, and consequences when copied (e.g., Tuesday May 4, 3, a 3-day stay at two hotels and one restaurant); and (2) substitute the place where the event is located (i.e., a city (Cleveland, Ohio) with a place familiar to the subjects living in the target countries. RESULTS: The event (fire) could be kept in all countries. The date had to be kept in all countries in the Netherlands because it corresponds to a commemoration (Remembrance of the Dead) and would have introduced a bias if kept. The verb (a “3 alarm fire”) was impossible to translate literally since no equivalent fire-classification exists in most target countries (except in Canada). It was decided to use synonyms of “big” to qualify “fire.” Syntax was also an issue especially in Korea, Japan, Romance and Germanic languages where the order of some segments had to be inverted. CONCLUSIONS: Although simple in its structure, the KBANS story memory was proved to be challenging to translate into 24 languages and required a rigorous methodology to preserve the intent of the original.

PRM110 STANDARDIZATION OF MENTAL HEALTH ASSESSMENT – USING ITEM RESPONSE THEORY (IRT) TO CROSS-CALIBRATE TWO SELF-REPORTED MENTAL HEALTH TOOLS: THE PATIENT HEALTH QUESTIONNAIRE (PHQ-9) AND THE SF-36V2 MENTAL HEALTH (MH) SCALE
Bjorner JB, White MK, Yarias AS
Optum, Lincoln, RI, USA
OBJECTIVES: Mental health can be measured by numerous instruments, but scores are usually not directly comparable. The heterogeneity of scale specific metrics seriously impairs comparability across study results and the communication among researchers and clinicians. We aimed to develop and evaluate methods for cross-calibration of the two popular mental health tools, the PHQ-9 and the SF-36V2 MH scale.

METHODS: We analyzed data from the United States and the UK including a general population sample (US: 216, UK: 355) and a sample with suspected depression (US: 467, UK: 98). We analyzed caregiver models tested whether the two instruments measured the same construct. Differential item function (DIF) between general population and depression samples was tested using logistic regression DIF tests. We estimated IFs using multiple imputed datasets (N=200) which were not adaptive to the training sample data, bias was low for all estimators. Estimators were evaluated by assessment of bias and variance. All four estimators were adaptive to the training sample data. For the two mapping algorithms and similar in the full sample estimators (bootstrap and cross-validation methods).

RESULTS: We applied three different methods for covariate matching to determine the “average effect of treatment on the treated” (ATT) of caregiving on mental health states (MH). The 1. Propensity score matched covariates within calipers; 2. Genic algorithm matching. RESULTS: All three methods provide adequate balance on the covariates used for matching. Methods 2 and 3 produce the best covariate balance, with absolute mean difference of less 0.0008 on all covariates and less than 0.0001 on the core set of covariates. Because methods that censor observations (i.e. matching within calipers) may artificially improve covariate balance, we take the ATT estimate from genetic matching to be the least biased estimate of the true effect. Using a standard 5-point self-report measure of mental health, caregivers, on average, report a mental health state that is 5.4% worse than non-caregivers (roughly one-fourth “less healthy” within any given scale range (e.g. 2-3, 3-4).

CONCLUSIONS: As all three methods provide adequate matching, our consideration turns to bias reduction and the fact that the genetic matching does not require that we estimate the propensity score prior to matching. We consider the drivers or caregiver MH and implications for health care policy.

PRM114 ARE INDUSTRY FUNDED NETWORK META-ANALYSES LOWER QUALITY? A PUBLICATION BIAS HETEROGENEITY AND ERROR ANALYSIS OF INDUSTRY FUNDED NETWORK META-ANALYSES COMPARED TO NON-INDUSTRY FUNDED META-ANALYSES
OBJECTIVES: To compare the quality and transparency of industry supported network meta-analyses with those with non-profit support or no support.

METHODS: We systematically searched OVID-Medline for network meta-analyses including at least one pharmacological. We reviewed each network meta-analysis and evaluated key general study characteristics, methodology, and transparency using a checklist of objective criteria derived from the ISPOR Taskforce’s recommendations for study conduct and reporting. We reported source of study funding when available. When source of funding was unclear or not reported we contacted the corresponding author. We compared the quality and transparency of industry supported network meta-analyses with those with non-profit support or no support.

RESULTS: Two hundred and fourteen studies met our inclusion criteria and were included in our dataset. Source of funding was identified for 211 studies (98.6%). Industry supported studies tended to be published in lower quality medical journals (p<0.01), and typically included fewer studies (p<0.05) and a smaller total number of patients (p<0.05). In terms of study transparency, industry supported studies less often reported the search terms (p<0.01) and, for analyses conducted using a Bayesian framework, presented the model code (p<0.01). Regarding study methodology, industry supported network meta-analyses less often reported a quality assessment of clinical studies included in the network meta-analysis (p<0.01), and less often compared the findings of traditional meta-analysis and network meta-analysis (p<0.01). With respect to presentation of findings, industry supported studies less often included a formal review of head-to-head comparisons (p<0.01), or provided a ranking of treatments (p<0.01).

CONCLUSIONS: We found that studies with non-profit support or no support funded tended to be more transparent and rigorous than industry funded network meta-analyses. Study findings emphasize that users of network meta-analyses should take great care to account for study quality when interpreting the findings of network meta-analyses.

PRM115 AUTOMATIC DEVELOPMENT OF CLINICAL PREDICTION MODELS WITH GENETIC PROGRAMMING: A CASE STUDY IN CARDIOVASCULAR DISEASE
Van der Giessen CA, Curtis CJ, France A, Spasic I
Geld University, Cardiff, UK
OBJECTIVES: To develop a computer program that automatically constructs a clinical prediction model from a dataset of patients with cardiovascular disease.

METHODS: We used genetic programming to construct a model of cardiovascular disease. The model was trained on a dataset of patients with cardiovascular disease and tested on a separate dataset of patients with cardiovascular disease. The model was evaluated using a receiver operating characteristic curve.

RESULTS: The model was able to accurately predict cardiovascular disease in a test dataset. The area under the receiver operating characteristic curve was 0.85, indicating excellent predictive accuracy.

CONCLUSIONS: Genetic programming is a powerful tool for constructing clinical prediction models. Further research is needed to evaluate the performance of genetic programming in other medical domains.